ITIS (THE INTEGRATED TAXONOMIC INFORMATION SYSTEM)

Courtney Ann Shaw
Vertebrate Zoology Librarian/Senior Reference Librarian
Smithsonian Institution Libraries
National Museum of Natural History
10th and Constitution Ave.
Washington, D. C. 20560 USA
E-mail: shawc@si.edu

ABSTRACT: World biological communities are currently using ITIS. Taxonomists are contributing to ITIS. "Mother" ITIS is based in the National Museum of Natural History (with other bases in Canada and Mexico). So, when I discovered that the 3rd and revised edition of *Using the Biological Literature* by Diane Schmidt (Marcel-Dekker, 2002) made no mention of ITIS (the Integrated Taxonomic Information System), I knew it was time to go on the road to explain this extremely useful database. I plan to talk about the history of ITIS, its relationship to other databases, fields, who uses it and how, including its use by librarians. I will talk about the Taxonomic Workbench as a way for scientists to enter data into the system, and how ITIS can be downloaded into others' databases to enhance their files. Taxonomists, oceanographers, ecologists, and natural history librarians are often looking for accurate scientific and common species names, synonyms, taxon authors, etc. They often use in-print and online sources. I believe ITIS is the place to start. It is a dynamic (includes revisions), integrated, accessible, well-documented database and databank. It is web-based, so it has broad public access. By integrating the research results of the world taxonomic community into a current coherent list of biological names, it standardizes taxonomic vocabulary. ITIS uses the traditional Linnean nomenclature system, (not a phylogenetic one) with a five kingdom hierarchical model—Protistans (currently protistan names are in ITIS in other kingdoms), Monera, Fungi, Plants, and Animals. Thus it includes multi-taxa at one site. For those areas designated in the Contributions section of the database, it is more comprehensive and accurate than other sources. ITIS strives for data quality and includes quality indicators for credibility, completeness, and currency, and also provides sources (experts, references, databases) for the names, often not present in other databases. i.e. scientist and date of when first the specimen (type) was named, and often links to bibliographic references which can give the original descriptions for Animalia (not required for Plantae) of the taxon. ITIS can be used at anywhere in the database management process: for collecting data/specimens, and cataloguing, querying, and archiving them. It can be incorporated into a bibliographic record, thus it helpful to librarians

as well. ITIS has four gateways; each somewhat different: U.S. (Hawaii node planned), North American, Canadian, Mexican and others are under consideration. At present it uses three languages: English, French, and Spanish, with Portuguese under development. ITIS is an evolving, ongoing effort. Because it cross-platforms- data can be shared, it can coalesce across time, geography, and organization. It is becoming the standard database for taxonomic information, and someday might become a one-stop shopper!

Part One

World biological communities are currently using ITIS. Taxonomists are contributing to ITIS. "Mother" ITIS is based in our museum. So, why not use ITIS as the starting point for taxonomic information?

Taxonomists and natural history librarians often look for accurate scientific and common species names, synonyms, taxon authors, etc. They frequently use in-print and online sources, e.g. BIOSIS (Index to Organism Names), Zoological Record (animals only), Biological Abstracts, CABI browse monographs. They search the Division's card files, visit the Natural History Museum, London or other online catalogs, NCBI's (National Center For Biotechnical Information) Taxonomic Browser, Pubmed (NIH-NCBI)) or Neave- Nomenclator zoologicus and other classic books, as well as specialized nomenclators and checklists, e.g., Simpson, Wilson and Reeder, Eschmeyer (an annotated nomenclature list).

ITIS is often the best first step in looking for this information. It is a dynamic (includes revisions), integrated, accessible, well-documented database and databank. It is web-based, with broad public access. By integrating the research results of the world taxonomic community into a current coherent list of biological names, it has a standardized vocabulary. ITIS uses the traditional Linnean nomenclature system, with five kingdom hierarchical model- Protistans (presently within other kingdoms), Monera, Fungi, Plants, and Animals, not a phylogenetic one. For those areas designated in the Contributions section, it is more comprehensive and accurate than other sources. It includes multi-taxa at one site. It has more elements/fields than many other databases, i.e. scientist and date of type specimen, links to bibliographic references (which can give original descriptions of the taxon). Fields that ITIS uses includes Scientific Names, Authors, Taxonomic Rank, Usage of Scientific Name, Reference for Scientific Name, Associated names (common names, synonymy, etc.), Linkages to hierarchical classification, Accepted name or Parent name. Additional Data Elements include the Taxonomic Serial Number (TSN), which is assigned to each scientific name record and is a unique, persistent identifier for names within the ITIS reference system. It provides a unique identifier (Taxonomic Serial Number). There are also Data Quality Indicators including Taxonomic Credibility Rating, which is a subjective rating on the perceived level of review and accuracy of taxonomic name and attributes, Taxonomic Completeness Rating (rank of genus and above) which reflects the extent of taxonomic

completeness for known species of a group (complete, partial, unknown), and Taxonomic Currency (rank of genus or above) the date on which a record is modified in online system (providing assistance to those downloading on a periodic basis), and Reference(s) which gives the year of taxonomic revision.

ITIS can be used anywhere in the database management process: collecting data/specimens, cataloguing, querying, and archiving. It can be incorporated into a bibliographic record.

ITIS is only the first step in searching for taxonomic information, since it contains only minimum data elements. Other sources have their own unique features, e.g. BIOSIS, the Index of Organism Names shows nomenclatural changes.

ITIS (U.S) includes information on North America and Surrounding Oceans (biota of importance to North America, including invasive and migratory species). The Database contains 1,750,000 World Species: Monera – 1,465, Protista – 15,509, Fungi – 3,857, Plantae – 85,411, and Animalia – 214,602.

ITIS has four gateways: U.S.(a Hawaiian node is in process), North American, Canadian, Mexican, and others are under consideration. ITIS (U.S.) is a centralized relational database manned by permanent staff who fundraise, contract, look for data, create partnerships, and are responsible for getting ITIS used. It is government coordinated. The offices are in this building (NHB) with DAVE NICOLSON (Head of Data Development Team), TODD McCONCHIE, JANET GOMON (Deputy Director), SI staff MICHAEL RUGGIERO (Director). USGS's ROY McDIARMID (Reptiles and Amphibians) was in at the core of ITIS's beginnings. Other experts include STEVE CAIRNS, IZ, and NOAA's DAVE HARDY, who is updating legacy data for fish.

ITIS (U. S.) is the data development site. One can download the whole database, query, send project scripts from machine to machine, and obtain reports. One can move the contents of the database to one's own application to develop a front-end system or cataloguing system. It links to Google images, Biobot (NBII) and Comments.

ITIS (Canada) SITI*ca – (Système d'Information Taxonomique Intégré) is sponsored by Agriculture and Agri-Food Canada (It is also in French). With its front-end emphasis for outreach to users, it has more links than the U.S. version: other sources, Google images, other Internet search portals, e.g. Lycos, and other specialized internet databases - GenBank, Medline (Pubmed), Agricola, Digital Library, etc. and into specimen collections via the BIOSC Gateway, i.e. the European Natural History Specimen Information Network (ENHSIN), The Species Analyst (TSA), and the World Information Network on Biodiversity (REMIB). ITIS/Canada has also added some competing classifications, e.g. Birds.

ITIS (Mexico) (Sistema Integrado de Información Taxonómica) SIIT*mx is under the leadership of CONABIO (LA COMISIÓN NACIONAL PARA EL CONOCIMIENTO Y

USO DE LA BIODIVERSIDAD) and is in Spanish. The North American link is the portal to the U.S., Canadian and Mexican Databases combined. ITIS North America has English, Spanish, and French language access, with Portuguese under development.

ITIS is evolving technically, and organizationally. My understanding is that it would like to head toward decentralization (now only U.S. ITIS inputs data). Eventually, ITIS would like users to be able to even edit online. Other future plans include talks with managers of SnoMed (American College of Pathologists), PubMed (NCBI)- which use their own taxonomic systems-about coordinating systems, and Zoological Record, where ZR would get new species information from them and in return, give them TSN, so that they could incorporate ITIS elements into their bibliographic record.

Someday ITIS might become a one-stop shop. It cross-platforms, so one can share data - across time, geography, and organization. It *is* becoming the standard databank/database for taxonomic information. ITIS is listed with other Databases in SIRIS. My suggestions to the reader are to place the ITIS web site on a personal toolbar.

ITIS URLS follow:
ITIS http://www.itis.usda.gov - (U.S.)
ITIS *NA http://sis.agr.gc.ca/itis/- (North American Portal)
ITIS *CA http://www.agr.gc.ca/itis - (Canada)
ITIS *MX http://siit.conabio.gob.mx - (Mexico)

For Additional Information Email:itiswebmaster@www.itis.usda.gov or gomon.janet@nmnh.si.edu

Part 2

History

ITIS evolved out of the National Oceanographic Data Center (NODC) Taxonomic Code, developed by the Department of Commerce, National Oceanic and Atmospheric Administration (NOAA), which wanted a coding system to link Linnean nomenclature with modern oceanographic data in a system for storage and retrieval. (Previous systems have been abbreviations, alphanumeric, or numeric schemes.) Dr. George Mueller, University of Alaska, derived the code from the VIMS (Virginia Institute of Marine Science) code published as a "Taxonomic Code for the Biota of the Chesapeake Bay." He created 10-digit, plus two for subspecies, digit couplets, each of which represents one or more levels of the taxonomic hierarchy. (NODC data originally loaded into ITIS are being actively reviewed and brought up to quality standards of ITIS. Dave Hardy is one of the many people reviewing that legacy data.)

The White House National Science and Technology Council's Committee on the Environment and Natural Resources identified the need for a consistent list of names of organisms as fundamental to the successful description and management of biological

resources and information. Their idea was to create an easily accessible database with reliable information on species names and their hierarchical classification. The group created a Memorandum of Understanding, Database Work Group (for database design and development), and a Taxonomy Work Group (for quality and integrity of the database). It was developed by Cooperation (Partnerships) among U. S. Federal Agencies: the U. S. Geological Survey, and the National Biological Survey (DOI), Environmental Protection Agency, Natural Resources Conservation Service and the Agriculture Research Service of the USDA, National Oceanic and Atmospheric Administration (DOC), and National Museum of Natural History (SI). A Data Development Team was established at the Smithsonian; focusing on adding/updating priority taxonomic groups. A public interface became available in 1996.

Since then ITIS has been joined by Agriculture and Agri-Foods Canada (1998) the National Park Service (1999) CONABIO (Mexico) (2001) and NatureServe (2002) The National Science Foundation PEET Program has provided supplemental funding to grantees for collaboration with ITIS. ITIS has worked with Species 2000 for the past two years to develop a "Catalog of Life", a comprehensive database of global species lists. As a founding partner in GBIF, ITIS is helping to develop GBIF's Electronic Catalog of Names of Known Organisms. ITIS is also supporting the Global Taxonomic Initiative with taxonomic information. ITIS' U.S. website, established in 1996, has additional portals for North America, Canada, Mexico, and Brazil, and provides interfaces in English, French, Spanish, and Portuguese.

Part 3

Scientists at universities, government agencies, and museums contribute their raw data, card files, publications, and individual databases to ITIS, and extrapolate information from the databank.

Scientists can input their own data into ITIS through the Taxonomic Workbench. The Taxonomic Workbench is a Windows-based (Microsoft Access), relational database which interfaces with the online system. Other software applications, such as word processors, spreadsheets, or databases can be programmed to align with ITIS elements and be uploaded as well.

To submit data to ITIS, scientists follow the submission guidelines, including how to submit their data, what data needs to be included, and in what format. They must be given a password to download the software and to build their own taxon lists. An example of this effort is the National Marine and Estuarine Invasion Database (SERC) that is maintained by Paul Fofonoff. Listings are then resubmitted to ITIS with the added data.

Portions of scientists' own taxonomies are being entered into ITIS, for instance, Alan Peterson's Zoonomen.com has taken Sibley/Monroe's "World List of Bird Names" and has added value to that list with the first bibliographic reference for the type species. ITIS

is absorbing it. *Mammal Species of the World*, and Eschmeyer's *Catalog of the Genera of Recent Fishes* are also being entered into ITIS.

Scientists may check to see if the species name in which they are interested is in ITIS. If they are looking for a few names a single search into the ITIS website is used. If many names are to be searched the "Compare Taxonomy/Nomenclature" tool is used. They also use ITIS to look up references outside their specific field.

Scientists act not only as data experts but as stewards (consultants). They review entries (the verification process) and know who the authorities are. These data developers submit files to a Database Manager for review, verification, and addition of missing data and quality control attributes. The Taxonomic Working Group (Janet Gomon, Roy McDiarmid, Mike Ruggiero) review the entries. A data load is then made (moving it to the online dat base), with quality control checks, assignment of a Taxonomic Serial Number, and a review by the submitter.

Scientists can use ITIS for collecting data on specimens. They can put data from the field into a "Palm Pilot," and download it at end of the day. This is especially useful when working on a large number of species at one time.

They may also use ITIS for querying. The Species Analyst (TSA) is a University of Kansas research project (which uses Extensible Mark-Up Language with a XML output option and the Dublin Core) may be accessed through ITIS (CA) using the BIOSC Gateway. The BIOSC harvester scans it (and other networks), retrieving their records. BIOSC names matching ITIS names are automatically cross-referenced and hyperlinked in both systems. When BIOSC records matching ITIS records are provided with geographic coordinates, they are automatically mappable, and a "Map It!" button is dynamically added to ITIS output page. Three botany databases, and FishNet, have been entered into SA.

Scientists and collection managers use ITIS to catalog collections. For instance, Biolink's Taxon Explorer (CSIRO-Commonwealth Scientific and Industrial Research Organisation) has classification, unomenclature, ecology, distribution, multimedia, and literature references. It has "find" and "favorites" features. The Specify Collections Management System (formerly MUSE) supports data entry, queries, and reports for single or multiple taxonomic discipline collections. It has an import function for taxonomic data that are formatted into ITIS Taxonomic Workbench database schema (tables.) It stores names and classification relationships in a taxonomic dictionary that can be used as an index and source of available names for your collection. It uses ITIS as an authority file.

Here at the Natural History Museum, The KE Emu Specimen Database (some records previously on SELGEM, then Inquire, now known as the Multimedia Catalogue), will have a dynamic link to ITIS as Name Server, so that people can use ITIS to find out what other names they need to use to search for their data. Presently, Invertebrates, Mineral

Sciences, Botany, Paleobiology, and Vertebrates have been entered; anthropological records will be entered next. At some point the database might feed into Species Analyst.

ITIS is being used in the web-driven Digital Libraries scientists help create. For instance, AMNH tagged their data with ITIS for The Congo Expedition 1909-1915. The Albatross Cruises (SIO Explorer-Scripps) which includes information on IZ and Fish, focuses on the Pacific, and will migrate into the Emu database. It will link specimen records, bibliographic databases, taxonomic databases, and digital images of type specimens. There will be full-text of important source literature. CalFlora (California Digital Library) also uses ITIS. The Biologia Centrali-Americana is a series of 63 volumes on the flora, fauna, and anthropology of Central America. It is the first biological survey of the area from Mexico to Panama, written in English by two explorers, Salvin and Godman. The Smithsonian Institution and five other institutions are collaborating on making full-text, images of text, and plates of the 58 volumes of biology. Scanning and rekeying the entire work has been contracted out. The XML Schema will make the documents fully searchable and provide links to specimen databases, such as Emu and Species Analyst; and taxonomic databases, such as ITIS via highlighted species name. Potentially it can link to the Alexandria Digital Library Gazetteer Server. Scanning is to be completed August 2004.

Nearctica.com and The Animal Diversity Web (University of Michigan, Department of Zoology) use ITIS as Authority Files. The integrated taxonomies/thesauri of SNOMED (Systematized Nomenclature of Medicine) of the College of American Pathologists, GEMET (General Multilingual Environmental Thesaurus); EIONET (European Environment Information and Observation Network), and the GCMD (Global Change Mastery Directory) all use ITIS for their metadata sources.

ITIS may be used for environmental inventorying and monitoring. Databases such as STORET (EPA), ECOTOX (Environmental Toxicity Database Trilogy), EMAP (Environmental Monitoring and Assessment Program), World Ocean Database, Coral Reef Information Management, USGS Breeding Bird Survey, and the National Park Species information System, are ITIS compliant and store the TSN in their databases. Other Organizations such as Intergovernmental Oceanographic Commission, Commission on Environmental Cooperation, North American Biodiversity Information Network (NABIN); Man and the Biosphere UNESCO)'s BRIM (the flora and fauna databases);and EcoPort (Ecology Knowledge Portal) query system of BioNET-INTERNATIONAL-Convention on Biodiversity, Global Network for Taxonomy, with locally owned and operated partnerships (LOOPS), do so as well.

Part 4

ITIS is important to librarians because it not only provides a common terminology between scientists and librarians, but also provides the current thinking of systematists on relationships amongst taxa. ITIS can be used for quick reference: finding accepted scientific names and synonyms (a thesaurus), vernacular names, first naming of an

animal, or where an animal falls in the taxonomic hierarchy, e.g. Family, Order, Genus. It also can be used as a spell-checker for scientific names, and to look up pictures of species, as well as their distribution.

By using ITIS to check the unfamiliar names (especially scientific ones) , catalogers can learn what an obscure book is about. Pictures in books don't always help! They can also apply ITIS's taxon names to bibliographic records. Catalogers can use ITIS to refine subject headings. They do this by first checking LCSH (Library of Congress Subject Headings) established headings (http://authorities.loc.gov/). If the subject heading is not there, librarians can submit names to be added as part of the Library of Congress's Program for Cooperative Cataloging (http://www.loc.gov/catdir/pcc/). The Subject Authority (SACO) page links to a list of Web resources for use in preparing SACO proposals. ITIS is on this list as a recognized authoritative source for scientific taxa.) Names to be added may be submitted to the Library of Congress subject thesaurus by any librarian. If an alternative term is found, the ITIS term maybe added as a cross reference.

ITIS is useful in determining where the genus and species fall within the L.C. system. It is more complete in terms of hierarchical levels because the L.C. authority files have Order and Families, but no Genus or species references. Library of Congress subject headings also sometimes retain old names and add new ones (which is confusing), such as Atelopodidae (currently not considered a family level category by many herpetologists), omit names, (e.g. ecaudata, but not the newer term for frogs, anura), and can misrepresent authorities, (i.e. Gans is an authority on Amphisbaenids, not all lizards).

Proposals for new LC classification numbers are also part of the SACO program. Resources such as ITIS can be used to propose new LC classification numbers.

An excellent example of the use of ITIS by libraries is *Linking Florida's Natural Heritage* (http://www.susdl.fcla.edu/lfnh) where museum collections (specimens) are linked to library collections (bibliographic literature) about specimens. The database links to ITIS for authority information and captures information in the 754 taxonomic field for MARC record enhancement. It is especially useful for common names.

An example of their use of ITIS is below:

MARC Enhancement

Partial MARC Record: field 754 with ITIS data

651: 4: |a Hillsborough |z 12057 |x FIPS
651: 4: |a Tampa Bay |z 3100206 |x HUC
752: : |a United States |b Florida |c Hillsborough |d Tampa Bay
754: : |a Kingdom: Animalia |a Phylum: Chordata |a Subphylum: Vertebrata |a Class: Chondrichthyes |a Subclass: Elasmobranchii |a Superorder: Selachimorpha |a Order: Carcharhiniformes |a Suborder: Scyliorhinoidei |a Family: Carcharhinidae |a Genus: Carcharhinus |a Species: Carcharhinus limbatus |a Author: (Valenciennes, 1839) |a Common Name: blacktip shark |a Synonym(s): Carcharhinus natator, Carcharias aethlorus, Carcharias ehrenbergi, Carcharias limbatus, Carcharias microps, Carcharias muelleri, Carcharias phorcys, Carcharias pleurotaenia, Galeolamna pleurotaenia tilsoni, Gymnorhinus abbreviatus, Gymnorrhinus abbreviatus |2 **TSN160318 Integrated Taxonomic Information System, 6/29/99**
773:0 : |t Fishery bulletin |g (1989) |g 87(4): |g 845-858

Effective access to bibliographic records using scientific information is through the use of proper terminology. Searchers will use the language that they know and understand. ITIS is the bridge that helps the cataloger/librarian reflect accurately the words/world of the scientist. Data may be in a traditional library catalog or embedded in the html of a web site, but without the proper terms the information is inaccessible.

References

Alliance News; The Newsletter of the Natural Science Collections Alliance. 2001. 29(6): December 2001.

Bank, David. 2002. All species great and small: taxonomists unite to catalog the planet's biodiversity; possible cost? $3 billion. *Wall Street Journal* January 22, 2002, s0 pB1 col 2.

Handwerk, Brian. 2002. Team races to catalog every species on Earth. [Online]. Available: http://news.nationalgeographic.com/news/2002/03/0305_0305_allspecies.html [Accessed 12/5/03].

Hardy, Jerry D. Jr. 1993. NODC taxonomic code links biology and computerized data processing. *Earth System Monitor* 4(2): 1-2,12.

McDiarmid, Roy. 1998. The integrated taxonomic information system. *Proceedings of the Taxonomic Authority files Workshop, Washington, D. C. June 22-23.* [Online]. Available: http://research.calacademy.org/taf/proceedings/itis.html [Accessed December 5, 2003].

Sugden, Andrew and Pennisi, Elizabeth. 2000. Diversity digitized. *Science.* 289: 2305-2314.

Taxonomy: all living things, online. 2002. *Nature* 418: 362-363.

Blum, Stanley. 2000. *An Overview of Biodiversity Informatics*. [Online]. Available: http://www.calacademy.org/research/informatics/sblum/pub/biodiv_informatics.html [Accessed December 5, 2003].

Acknowledgements

- Dr. Roy McDiarmid, U.S. Geological Survey, Research Zoologist, Patuxent Wildlife Research Center (Division of Amphibians and Reptiles)
- Janet Gomon, Deputy Director, ITIS
- Suzanne Pilsk, Cataloger, Smithsonian Institution Libraries
- Lowell Ashley, Cataloger, Smithsonian Institution Libraries
- Adam Schiff, Cataloger, University of Washington, Seattle
- Stephanie Haas, Librarian, University of Florida, Gainesville.
- Dr. Paul Fofonoff, Biologist, SERC
- Dr.Scott Miller, Chairman, Dept. of Systematic Biology